Journal of Modern Applied Statistical Methods

Volume 17 | Issue 1 Article 14

6-20-2018

Regressions Regularized by Correlations

Stan Lipovetsky
GfK North America, stan.lipovetsky@gfk.com

Regressions Regularized by Correlations

Cover Page Footnote
Acknowledgements: I am thankful to two reviewers for their comments helped to improve the paper.

Journal of Modern Applied Statistical Methods
May 2018, Vol. 17, No. 1, eP2530
doi: 10.22237/jmasm/1529506907

בס"ד

ISSN 1538 − 9472

Regressions Regularized by Correlations

Stan Lipovetsky
GfK North America
Minneapolis, MN

The regularization of multiple regression by proportionality to correlations of predictors with dependent variable is applied to the least squares objective and normal equations to relax the exact equalities and to get a robust solution. This technique produces models not prone to multicollinearity and is very useful in practical applications.

Keywords: Multiple regression, multicollinearity, regularizations, robust solutions

Introduction

Regression modeling is widely used in statistical analysis of data. Ordinary least squares (OLS) models are efficient for prediction but often demonstrate poor results in the analysis of individual predictors because of the multicollinearity effects. Multicollinearity makes the parameter estimates fluctuate wildly with a negligible change in the sample, causes reduction in statistical power, and leads to wider confidence intervals for the coefficients so they could be incorrectly identified as being insignificant (Grapentine, 1997; Mason & Perreault, 1991). To overcome these deficiencies, the ridge regression (RR) and its modifications have been developed with a quadratic L_2-metric regularization imposed on the parameters (Hoerl & Kennard, 1970, 2000; Golub, Heath, & Wahba, 1979; Hawkins & Yin, 2002; Liu & Gao, 2011; Hansen, 2016). Regularization based on the linear absolute L_1-metric in the lasso regression and the linear and quadratic metrics combined in the elastic net or sparse analysis are also known (Tibshirani, 1996; Hastie, Tibshirani, & Friedman, 2001; Efron, Hastie, Johnstone, & Tibshirani, 2004). Various other penalizing and constraining methods have been developed as well (for instance, Lipovetsky & Conklin, 2005, 2015; Lipovetsky, 2010, 2013; and the references within).

doi: 10.22237/jmasm/1529506907 | Accepted: February 16, 2017; Published: June 20, 2018.
Correspondence: Stan Lipovetsky, stan.lipovetsky@gfk.com

The purpose of this study is to propose a new regularization approach based on the criterion of regression coefficients proportionality to the pair correlations of predictors with dependent variable. Such a penalizing condition can be added to the OLS objective for building regression, so it may be called the regularized OLS (ROLS). Otherwise, it can be applied to the system of normal equations (NE) for parameters estimation in order to relax the exact equalities and to get a solution aligned by the correlation structure; we can call it the regularized normal equations (RNE). In contrast to ridge, lasso, and elastic net models, which depend on the profiling parameter, the ROLS and RNE solutions are uniquely defined only by the correlations among the variables so they do not require additional estimation of the penalizing parameter.

For numerical comparison with stable non-prone to multicollinearity results, the Shapley value regression (SVR) is used as a benchmark model. Shapley value is a construct from cooperative game theory used to evaluate the worth of participants over all possible combinations of them, and it can be employed for building regression models in a nonlinear estimation (Shapley, 1953; Roth, 1988; Lipovetsky & Conklin, 2001).

Ordinary Least Squares and Ridge Regressions

Consider briefly some OLS relations needed for further analysis. For the standardized variables, a multiple linear regression $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_n x_{in} + \varepsilon$ can be written in matrix form

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{1}$$

where $\mathbf{X}$ is an m by n matrix with elements x_{ij} of i-th observations ($i = 1,\ldots, m$) by j-th independent variables ($j = 1,\ldots, n$), $\mathbf{y}$ is the vector of observations for the dependent variable, $\boldsymbol{\beta}$ is the n-th order vector of beta-coefficients for the standardized regression, and $\boldsymbol{\varepsilon}$ is a vector of deviations from the theoretical model. OLS objective minimizes the sum of squared deviations:

$$\begin{aligned} S^2 &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2 = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}'\mathbf{y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \\ &= 1 - 2\boldsymbol{\beta}'\mathbf{r} + \boldsymbol{\beta}'\mathbf{C}\boldsymbol{\beta} \to \min \end{aligned} \tag{2}$$

where prime denotes transposition, variance of the standardized **y** equals one, and **C** and **r** denote correlation matrix of regressors of *x* and vector of correlations of **y** with them, respectively:

$$\mathbf{y}'\mathbf{y} = 1, \quad \mathbf{C} = \mathbf{X}'\mathbf{X}, \quad \mathbf{r} = \mathbf{X}'\mathbf{y} \tag{3}$$

The first order condition $dS^2 / d\boldsymbol{\beta} = 0$ of the objective (2) minimized by the vector $\boldsymbol{\beta}$ yields the so called system of normal equations (NE)

$$\mathbf{C}\boldsymbol{\beta} = \mathbf{r} \tag{4}$$

The solution of NE produces a vector of standardized beta-coefficients of regression (1):

$$\boldsymbol{\beta} = \mathbf{C}^{-1}\mathbf{r} \tag{5}$$

The quality of the model is estimated by the residual sum of squares (2), or by the coefficient of multiple determination defined from (2) as

$$R^2 = 1 - S^2 = 2\boldsymbol{\beta}'\mathbf{r} - \boldsymbol{\beta}'\mathbf{C}\boldsymbol{\beta} \tag{6}$$

The minimum value S^2 (2) corresponds to the maximum value of R^2 (6), so with (4) and (5) the coefficient of multiple determination for the OLS solution can be presented in the equivalent forms:

$$R^2 = \boldsymbol{\beta}'\mathbf{r} = \boldsymbol{\beta}'\mathbf{C}\boldsymbol{\beta} = \mathbf{r}'\mathbf{C}^{-1}\mathbf{r} \tag{7}$$

If predictors are highly correlated or multicollinear, the matrix **C** becomes ill-conditioned (its determinant is close to zero) and its inverted matrix in (5) produces a solution with highly inflated coefficients, with values which can have signs opposite to the signs of pair correlations of *x*s with **y**, and presumably important variables become statistically insignificant. The model can be applied for prediction, but it becomes practically useless for the analysis and interpretation of the predictors' role in the model.

A well-known tool for overcoming difficulties of multicollinearity is suggested by ridge regression (RR). Adding to the OLS objective (2) a penalizing function of the squared norm of regression coefficients for preventing their inflation leads to the conditional objective:

$$S^2 = \left\| \mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\mathrm{RR}} \right\|^2 + k \left\| \boldsymbol{\beta}_{\mathrm{RR}} \right\|^2 = 1 - 2\boldsymbol{\beta}'_{\mathrm{RR}}\mathbf{r} + \boldsymbol{\beta}'_{\mathrm{RR}}\mathbf{C}\boldsymbol{\beta}_{\mathrm{RR}} + k\boldsymbol{\beta}'_{\mathrm{RR}}\boldsymbol{\beta}_{\mathrm{RR}} \rightarrow \min \tag{8}$$

where $\boldsymbol{\beta}_{\mathrm{RR}}$ denotes a vector of ridge estimates for the coefficients in the model (1) and k is a positive parameter. Minimizing (8) subject to the vector $\boldsymbol{\beta}_{\mathrm{RR}}$ yields the system of equations

$$\left(\mathbf{C} + k\mathbf{I}\right)\boldsymbol{\beta}_{\mathrm{RR}} = \mathbf{r} \tag{9}$$

where $\mathbf{I}$ is the n-th order identity matrix. The solution of this system is

$$\boldsymbol{\beta}_{\mathrm{RR}} = \left(\mathbf{C} + k\mathbf{I}\right)^{-1}\mathbf{r} \tag{10}$$

It is the vector of RR parameters, and it exists even for a singular matrix $\mathbf{C}$. Using the general expression (6) with the solution (10) produces the coefficient of multiple determination for RR:

$$\begin{aligned} R^2_{\mathrm{RR}} &= 2\boldsymbol{\beta}'_{\mathrm{RR}}\mathbf{r} - \boldsymbol{\beta}'_{\mathrm{RR}}\mathbf{C}\boldsymbol{\beta}_{\mathrm{RR}} = 2\mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{r} - \mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{C}\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{r} \\ &= 2\mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{r} - \mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\left[\left(\mathbf{C}+k\mathbf{I}\right) - k\mathbf{I}\right]\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{r} \\ &= \mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-1}\mathbf{r} + k\cdot\mathbf{r}'\left(\mathbf{C}+k\mathbf{I}\right)^{-2}\mathbf{r} = \boldsymbol{\beta}'_{\mathrm{RR}}\mathbf{r} + k\boldsymbol{\beta}'_{\mathrm{RR}}\boldsymbol{\beta}_{\mathrm{RR}} \end{aligned} \tag{11}$$

With k reaching zero, the RR solution (9)-(11) reduces to OLS regression (4)-(7). A value for the ridge parameter k can be estimated using cross-validation (Golub et al., 1979), but it could depend on the aim of the modeling. For instance, the ridge solution can be profiled by k in order to choose a vector with the parameters of the same signs as the pair correlations of xs with $\mathbf{y}$, which can be required by the content of a problem and facilitate interpretability of the individual predictors (Lipovetsky, 2010).

Regularization of OLS by Proportionality to Pair Correlations

Suppose we are interested in building a model with parameters proportional to the pair correlations $\mathbf{r}$ of the dependent variable with predictors (those are also the coefficients of paired regressions of $\mathbf{y}$ by each x separately). Then in place of parameters constraint against inflation in RR technique (8), let us consider such regularization for regression:

$$\begin{aligned} S^2 &= \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\mathrm{ROLS}}\|^2 + \|\boldsymbol{\beta}_{\mathrm{ROLS}} - k\mathbf{r}\|^2 \\ &= 1 - 2\boldsymbol{\beta}'_{\mathrm{ROLS}}\mathbf{r} + \boldsymbol{\beta}'_{\mathrm{ROLS}}\mathbf{C}\boldsymbol{\beta}_{\mathrm{ROLS}} + \boldsymbol{\beta}'_{\mathrm{ROLS}}\boldsymbol{\beta}_{\mathrm{ROLS}} - 2k\boldsymbol{\beta}'_{\mathrm{ROLS}}\mathbf{r} + k^2\mathbf{r}'\mathbf{r} \rightarrow \min \end{aligned} \tag{12}$$

where the vector $\boldsymbol{\beta}_{\mathrm{ROLS}}$ denotes the regularized OLS estimation ROLS of coefficients in the model (1). Taking derivatives by the vector $\boldsymbol{\beta}'_{\mathrm{ROLS}}$ and by parameter k and setting them equal to zero yields the system for minimization of the objective (12):

$$\mathbf{C}\boldsymbol{\beta}_{\mathrm{ROLS}} - \mathbf{r} + \boldsymbol{\beta}_{\mathrm{ROLS}} - k\mathbf{r} = 0, \quad k\mathbf{r}'\mathbf{r} - \boldsymbol{\beta}'_{\mathrm{ROLS}}\mathbf{r} = 0 \tag{13}$$

From the second equation in (13) we get the parameter $k = \boldsymbol{\beta}'_{\mathrm{ROLS}}\mathbf{r}/\mathbf{r}'\mathbf{r}$ and substitute it into the first equation in (13), producing the following matrix equation:

$$\left(\mathbf{C} + \mathbf{I} - \frac{1}{\mathbf{r}'\mathbf{r}}\mathbf{r}\mathbf{r}'\right)\boldsymbol{\beta}_{\mathrm{ROLS}} = \mathbf{r} \tag{14}$$

where $\mathbf{I}$ is the identity matrix, $\mathbf{rr}'$ is the matrix of the outer product of the vector of correlations by itself, and $\mathbf{r}'\mathbf{r}$ is the scalar product. Inverting the matrix in (14) yields the solution:

$$\boldsymbol{\beta}_{\mathrm{ROLS}} = \left(\mathbf{C} + \mathbf{I} - \frac{1}{\mathbf{r}'\mathbf{r}}\mathbf{r}\mathbf{r}'\right)^{-1}\mathbf{r} \tag{15}$$

This ROLS solution does not depend on the parameter k but is uniquely defined by the correlation matrix $\mathbf{C}$ and vector $\mathbf{r}$.

The solution (15) can be simplified as follows: Using notations

$$\mathbf{A} = \mathbf{C} + \mathbf{I}, \quad \boldsymbol{\rho} = \frac{1}{\sqrt{\mathbf{r}'\mathbf{r}}}\mathbf{r} \tag{16}$$

and applying the known Sherman-Morrison formula for matrix inversion

$$\left(\mathbf{A} - \boldsymbol{\rho\rho}'\right)^{-1} = \mathbf{A}^{-1} + \frac{\mathbf{A}^{-1}\boldsymbol{\rho\rho}'\mathbf{A}^{-1}}{1 - \boldsymbol{\rho}'\mathbf{A}^{-1}\boldsymbol{\rho}} \tag{17}$$

we can transform (15) to the form

$$\begin{aligned} \boldsymbol{\beta}_{\text{ROLS}} &= \mathbf{A}^{-1}\mathbf{r} + \frac{\mathbf{A}^{-1}\boldsymbol{\rho} \cdot \boldsymbol{\rho}'\mathbf{A}^{-1}\mathbf{r}}{1 - \boldsymbol{\rho}'\mathbf{A}^{-1}\boldsymbol{\rho}} = \mathbf{A}^{-1}\boldsymbol{\rho}\sqrt{\mathbf{r}'\mathbf{r}} + \frac{\boldsymbol{\rho}'\mathbf{A}^{-1}\boldsymbol{\rho}}{1 - \boldsymbol{\rho}'\mathbf{A}^{-1}\boldsymbol{\rho}}\mathbf{A}^{-1}\boldsymbol{\rho}\sqrt{\mathbf{r}'\mathbf{r}} \\ &= \frac{1}{1 - \boldsymbol{\rho}'\mathbf{A}^{-1}\boldsymbol{\rho}}\mathbf{A}^{-1}\boldsymbol{\rho}\sqrt{\mathbf{r}'\mathbf{r}} = \frac{\mathbf{r}'\mathbf{r}}{\mathbf{r}'\mathbf{r} - \mathbf{r}'\mathbf{A}^{-1}\mathbf{r}}\mathbf{A}^{-1}\mathbf{r} \\ &= \frac{\mathbf{r}'\mathbf{r}}{\mathbf{r}'\left[\mathbf{I} - \left(\mathbf{C} + \mathbf{I}\right)^{-1}\right]\mathbf{r}}\left(\mathbf{C} + \mathbf{I}\right)^{-1}\mathbf{r} \end{aligned} \tag{18}$$

Also, applying the transformation

$$\mathbf{C}\left(\mathbf{C} + \mathbf{I}\right)^{-1} = \left[\left(\mathbf{C} + \mathbf{I}\right) - \mathbf{I}\right]\left(\mathbf{C} + \mathbf{I}\right)^{-1} = \mathbf{I} - \left(\mathbf{C} + \mathbf{I}\right)^{-1} \tag{19}$$

the expression (18) can be represented as follows:

$$\boldsymbol{\beta}_{\text{ROLS}} = \frac{\mathbf{r}'\mathbf{r}}{\mathbf{r}'\mathbf{C}\left(\mathbf{C} + \mathbf{I}\right)^{-1}\mathbf{r}}\left(\mathbf{C} + \mathbf{I}\right)^{-1}\mathbf{r} \tag{20}$$

The result (20) means that the ROLS solution is proportional to the RR regression (10) for $k = 1$ with the quotient of two quadratic forms $\mathbf{r'r}$ and $\mathbf{r'C}(\mathbf{C} + \mathbf{I})^{-1}\mathbf{r}$. The ROLS criterion (12) can be combined with the RR criterion (8) in one conditional objective

$$S^2 = \left\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}_{\text{RR.ROLS}}\right\|^2 + k_1\left\|\boldsymbol{\beta}_{\text{RR.ROLS}} - k_2\mathbf{r}\right\|^2 \to \min \tag{21}$$

It is easy to show that its solution is reduced to a problem similar to (9), namely $(\mathbf{C} + k_1\mathbf{I})\boldsymbol{\beta}_{\text{RR.ROLS}} = k_3\mathbf{r}$ with the parameter $k_3 = 1 + k_1k_2$. This problem defines the two-parameter ridge regression model considered in Lipovetsky and Conklin (2005).

Regularization of Normal Equations

Penalizing by the proportionality of model coefficients to the pair correlation structure (12) can be applied to the normal equations (4) as well. Indeed, any change of OLS coefficients corresponds to deviations from the exact relations (4) from which these coefficients were estimated. Search for parameters aligned by structure of pair correlations of *x*s with **y** can only mean that the relations (4) could be fulfilled approximately. Thus, instead of exact equations, the relations (4) can be considered in their squared norm minimization, subject to the proportionality regularization as follows:

$$\begin{aligned} S^2 &= \left\|\mathbf{r} - \mathbf{C}\boldsymbol{\beta}_{\text{RNE}}\right\|^2 + \left\|\boldsymbol{\beta}_{\text{RNE}} - k\cdot\mathbf{r}\right\|^2 \\ &= \mathbf{r}'\mathbf{r} - 2\boldsymbol{\beta}'_{\text{RNE}}\mathbf{C}\mathbf{r} + \boldsymbol{\beta}'_{\text{RNE}}\mathbf{C}^2\boldsymbol{\beta}_{\text{RNE}} + \boldsymbol{\beta}'_{\text{RNE}}\boldsymbol{\beta}_{\text{RNE}} - 2k\boldsymbol{\beta}'_{\text{RNE}}\mathbf{r} + k^2\mathbf{r}'\mathbf{r} \to \min \end{aligned} \tag{22}$$

where the vector $\boldsymbol{\beta}_{\text{RNE}}$ denotes the regularized normal equations estimator RNE for coefficients in the model (1).

Similarly to minimization in (12), from derivatives of the objective (22) by the vector $\boldsymbol{\beta}'_{\text{RNE}}$ and by parameter *k*, we obtain the system of equations

$$\mathbf{C}^2\boldsymbol{\beta}_{\text{RNE}} - \mathbf{C}\mathbf{r} + \boldsymbol{\beta}_{\text{RNE}} - k\mathbf{r} = 0, \quad \mathbf{k}\mathbf{r}'\mathbf{r} - \boldsymbol{\beta}'_{\text{RNE}}\mathbf{r} = 0 \tag{23}$$

In the same way as we obtained (14) from (13), the second equation (23) gives $k = \boldsymbol{\beta}'_{\text{RNE}}\mathbf{r}/\mathbf{r}'\mathbf{r}$, and substituting it into the first equation (23) yields the matrix equation

$$\left(\mathbf{C}^2 + \mathbf{I} - \frac{1}{\mathbf{r}'\mathbf{r}}\mathbf{r}\mathbf{r}'\right)\boldsymbol{\beta}_{\text{RNE}} = \mathbf{C}\mathbf{r} \tag{24}$$

Thus, the solution for regularized normal equations is

$$\boldsymbol{\beta}_{\mathrm{RNE}} = \left(\mathbf{C}^2 + \mathbf{I} - \frac{1}{\mathbf{r}'\mathbf{r}} \mathbf{r}\mathbf{r}' \right) \mathbf{C}\mathbf{r} \tag{25}$$

These RNE coefficients also do not depend on the parameter k and are directly defined by the correlation matrix **C** and vector **r**.

It is interesting to note that in minimization of deviations $\mathbf{r} - \mathbf{C}\boldsymbol{\beta}_{\mathrm{RNE}}$ in (22), the vector **r** and matrix **C** can be seen as analogues of the vector of dependent variable and matrix of predictors in the objectives (2) and (12). Instead of matrix $\mathbf{C} = \mathbf{X}'\mathbf{X}$ and vector $\mathbf{r} = \mathbf{X}'\mathbf{y}$ (3), now $\mathbf{C}^2 = \mathbf{C}'\mathbf{C}$ and $\mathbf{C}\mathbf{r}$, respectively. With such a replacement, the ROLS solution (14)-(15) can be transformed into the RNE results (24)-(25). Applying (16)-(17) with a new matrix $\mathbf{A} = \mathbf{C}^2 + \mathbf{I}$ to the formula (25) gives

$$\boldsymbol{\beta}_{\mathrm{RNE}} = \left(\mathbf{C} + \frac{\mathbf{r}'\mathbf{C}\left(\mathbf{C}^2 + \mathbf{I}\right)^{-1}\mathbf{r}}{\mathbf{r}'\mathbf{C}^2\left(\mathbf{C}^2 + \mathbf{I}\right)^{-1}\mathbf{r}} \mathbf{I} \right) \left(\mathbf{C}^2 + \mathbf{I}\right)^{-1} \mathbf{r} \tag{26}$$

which is more complicated than the solution (20). With the general expression (6) we can construct the coefficient of multiple determination for RNE (26) as well.

Additional Linear Adjustment

Any regularized OLS solution can be proportionally adjusted for improving the total quality of the model fit. Denoting by **b** a vector of ridge coefficients $\boldsymbol{\beta}_{\mathrm{RR}}$ (10), regularized OLS solution $\boldsymbol{\beta}_{\mathrm{ROLS}}$ (20), or regularized normal equations solution $\boldsymbol{\beta}_{\mathrm{RNE}}$ (26), consider such an additional adjustment to a new vector $\mathbf{b}_{\mathrm{adj}}$ with the unknown term q:

$$\mathbf{b}_{\mathrm{adj}} = q\mathbf{b} \tag{27}$$

Using the expression (6) for quality of fit with adjusted solution (27) yields a quadratic function in q:

$$R^2_{\mathrm{adj}} = 2q\mathbf{b}'\mathbf{r} - q^2\mathbf{b}'\mathbf{C}\mathbf{b} \tag{28}$$

The value q for which this concave function reaches its maximum is

$$q = \frac{\mathbf{b}'\mathbf{r}}{\mathbf{b}'\mathbf{C}\mathbf{b}} \tag{29}$$

and the maximum of coefficient of multiple determination for adjusted solution equals

$$R^2_{\text{adj}} = \frac{(\mathbf{b}'\mathbf{r})^2}{\mathbf{b}'\mathbf{C}\mathbf{b}} \tag{30}$$

If $\mathbf{b} = \boldsymbol{\beta}$ in OLS solution (5), $q = 1$ and R^2_{adj} (30) coincides with the R^2 in OLS (7). But for any other solution $\mathbf{b}$ in (10), (20), or (26), the adjustment (27) changes the solution by the term (29):

$$\mathbf{b}_{\text{adj}} = \frac{\mathbf{b}'\mathbf{r}}{\mathbf{b}'\mathbf{C}\mathbf{b}}\mathbf{b} \tag{31}$$

For example, in the ridge model $\mathbf{b} = \boldsymbol{\beta}$ of solution (10) the adjustment term (29) is

$$q_{\text{RR}} = \frac{\mathbf{r}'(\mathbf{C}+k\mathbf{I})^{-1}\mathbf{r}}{\mathbf{r}'\mathbf{C}(\mathbf{C}+k\mathbf{I})^{-2}\mathbf{r}} \tag{32}$$

Then the ridge solution (10) adjusted by the term (32) becomes

$$\boldsymbol{\beta}_{\text{RR.adj}} = \frac{\mathbf{r}'(\mathbf{C}+k\mathbf{I})^{-1}\mathbf{r}}{\mathbf{r}'\mathbf{C}(\mathbf{C}+k\mathbf{I})^{-2}\mathbf{r}}(\mathbf{C}+\mathbf{I})^{-1}\mathbf{r} \tag{33}$$

and the corresponding quality of the fit (30) equals

$$R^2_{\text{RR.adj}} = \frac{\left[\mathbf{r}'(\mathbf{C}+k\mathbf{I})^{-1}\mathbf{r}\right]^2}{\mathbf{r}'\mathbf{C}(\mathbf{C}+k\mathbf{I})^{-2}\mathbf{r}} \tag{34}$$

The ROLS solution (20) is proportional to the ridge model (10) with $k = 1$, so the ROLS adjusted solution is

$$\boldsymbol{\beta}_{\text{ROLS.adj}} = \frac{\mathbf{r}'(\mathbf{C}+\mathbf{I})^{-1}\mathbf{r}}{\mathbf{r}'\mathbf{C}(\mathbf{C}+\mathbf{I})^{-2}\mathbf{r}}(\mathbf{C}+\mathbf{I})^{-1}\mathbf{r} \qquad (35)$$

and the corresponding quality of the fit equals

$$R^2_{\text{ROLS.adj}} = \frac{\left[\mathbf{r}'(\mathbf{C}+\mathbf{I})^{-1}\mathbf{r}\right]^2}{\mathbf{r}'\mathbf{C}(\mathbf{C}+\mathbf{I})^{-2}\mathbf{r}} \qquad (36)$$

The relations (33)-(34) for $k = 1$ reduce to (35)-(36). Similarly, the adjustment can be performed for RNE solution (26) as well. Analytical formulae can be bulky, but there is no problem in numerical adjustment estimation (27)-(30).

The adjusted solutions have an important feature which makes them similar to the OLS solution in the following aspect: As we see in (7), there is equality $\boldsymbol{\beta}'\mathbf{r} = \boldsymbol{\beta}'\mathbf{C}\boldsymbol{\beta}$ for the OLS solution, but it does not hold for other solutions, $\mathbf{b}'\mathbf{r} \neq \mathbf{b}'\mathbf{C}\mathbf{b}$. However, for adjusted solutions this equality is true. Indeed, using (31) find the scaler product

$$\mathbf{b}'_{\text{adj}}\mathbf{r} = \frac{\mathbf{b}'\mathbf{r}}{\mathbf{b}'\mathbf{C}\mathbf{b}}\mathbf{b}'\mathbf{r} = \frac{(\mathbf{b}'\mathbf{r})^2}{\mathbf{b}'\mathbf{C}\mathbf{b}} = R^2_{\text{adj}} \qquad (37)$$

and the quadratic form

$$\mathbf{b}'_{\text{adj}}\mathbf{C}\mathbf{b}_{\text{adj}} = \left(\frac{\mathbf{b}'\mathbf{r}}{\mathbf{b}'\mathbf{C}\mathbf{b}}\right)^2\mathbf{b}'\mathbf{C}\mathbf{b} = \frac{(\mathbf{b}'\mathbf{r})^2}{\mathbf{b}'\mathbf{C}\mathbf{b}} = R^2_{\text{adj}} \qquad (38)$$

Both expressions (37)-(38) reduce to the same one in (30), so

$$\mathbf{b}'_{\text{adj}}\mathbf{r} = \mathbf{b}'_{\text{adj}}\mathbf{C}\mathbf{b}_{\text{adj}} \qquad (39)$$

which means that, for an adjusted solution, the term q (29) would be equal to one or an adjusted solution cannot be further improved.

Using the relations (3), rewrite (39) as $\mathbf{b}'_{\text{adj}}\mathbf{X}'\mathbf{y} = \mathbf{b}'_{\text{adj}}\mathbf{X}'\mathbf{X}\mathbf{b}_{\text{adj}}$, and let $\tilde{\mathbf{y}} = \mathbf{X}\mathbf{b}_{\text{adj}}$ denote the theoretical values of the dependent variable predicted by the model; then

(39) can be represented as $\tilde{\mathbf{y}}'\mathbf{y} = \tilde{\mathbf{y}}'\tilde{\mathbf{y}}$. This equality is known for the OLS model, but it also holds for the considered adjusted solutions. This equality also shows

$$\tilde{\mathbf{y}}'\left(\mathbf{y} - \tilde{\mathbf{y}}\right) = \tilde{\mathbf{y}}'\boldsymbol{\varepsilon} = \mathbf{b}'_{\text{adj}}\mathbf{X}'\boldsymbol{\varepsilon} = 0 \tag{40}$$

so the theoretical vector $\tilde{\mathbf{y}} = \mathbf{X}\mathbf{b}_{\text{adj}}$, being a linear combination of columns $\mathbf{x}_k$ in the design matrix $\mathbf{X}$, is not correlated with the vector of errors in (1). More detail on the adjustment with any solution, including estimation of bias, efficiency, cross-validation of the estimated parameters, in and out-of-sample forecasts, and other characteristics of model quality, are given in Lipovetsky and Conklin (2005, 2015) and Lipovetsky (2010, 2013).

Numerical Results

Consider an example from a real marketing research project for a pharmaceutical company on a cold sore healthcare product. The purchase interest as the dependent variable and 35 attributes as predictors were measures in a 10-point Likert scale, and data were gathered from 1023 respondents – more details are given in Lipovetsky and Conklin (2015). The aim of modeling was to measure the input of the predictors in their influence on the dependent variable, and to compare the results of SVR and ridge models with the newly suggested techniques of ROLS and RNE. Table 1 presents in columns the vector of pair correlations **r** and beta-coefficients of several regression models: OLS (5), SVR and its adjustment by (31), RR (10) for $k = 1$ and its adjustment (33), ROLS (20) with its adjustment (35), and RNE (26) with its adjustment too. The bottom lines contain the coefficients of multiple determination R^2 and the adjustment parameter q (29).

All correlations in Table 1 are positive, but because of multicollinearity more than a third of the predictors (13 out of 35) receive negative signs in the OLS regression. However, the regularized models have all positive parameters as it is expected by the predictors meaning and good R^2. The parameter q is very close to one for SVR and two new regularized models, so they practically coincide with their adjusted versions, but q is higher for the ridge regression so solution RR_{adj} differs from RR and its quality improves with adjustment.

Table 1. Correlations of *x*s with **y** and comparison of several regression models

Variables	r	OLS	SVR	SVR_{adj}	RR	RR_{adj}	ROLS	$ROLS_{adj}$	RNE	RNE_{adj}
x1	0.698	0.079	0.038	0.038	0.043	0.045	0.045	0.045	0.040	0.040
x2	0.584	-0.038	0.024	0.024	0.010	0.010	0.010	0.010	0.018	0.019
x3	0.689	0.017	0.032	0.032	0.032	0.034	0.034	0.034	0.037	0.037
x4	0.698	-0.004	0.034	0.034	0.033	0.035	0.034	0.035	0.041	0.041
x5	0.716	0.076	0.040	0.040	0.051	0.054	0.053	0.054	0.049	0.050
x6	0.675	-0.039	0.031	0.031	0.020	0.021	0.021	0.021	0.032	0.033
x7	0.674	0.055	0.029	0.029	0.036	0.038	0.038	0.038	0.037	0.037
x8	0.720	0.084	0.040	0.040	0.050	0.053	0.053	0.053	0.049	0.049
x9	0.715	0.068	0.036	0.036	0.041	0.043	0.043	0.043	0.040	0.040
x10	0.650	-0.027	0.030	0.030	0.018	0.019	0.019	0.019	0.030	0.030
x11	0.611	0.030	0.029	0.029	0.029	0.031	0.030	0.031	0.032	0.032
x12	0.734	0.147	0.044	0.044	0.058	0.061	0.060	0.061	0.048	0.048
x13	0.569	0.039	0.025	0.025	0.027	0.029	0.029	0.029	0.026	0.026
x14	0.613	0.004	0.027	0.027	0.018	0.019	0.019	0.019	0.020	0.021
x15	0.634	0.105	0.034	0.034	0.041	0.043	0.043	0.043	0.030	0.031
x16	0.673	-0.029	0.029	0.029	0.017	0.018	0.018	0.018	0.029	0.029
x17	0.423	-0.017	0.014	0.014	0.003	0.003	0.003	0.003	0.002	0.003
x18	0.570	0.022	0.022	0.022	0.020	0.021	0.021	0.021	0.021	0.021
x19	0.542	0.041	0.023	0.023	0.018	0.020	0.019	0.020	0.014	0.014
x20	0.701	-0.013	0.036	0.036	0.033	0.035	0.035	0.035	0.041	0.042
x21	0.486	-0.046	0.019	0.019	0.005	0.005	0.005	0.005	0.012	0.012
x22	0.525	-0.010	0.019	0.019	0.013	0.013	0.013	0.013	0.016	0.016
x23	0.570	0.027	0.024	0.024	0.022	0.023	0.023	0.023	0.021	0.021
x24	0.626	-0.024	0.027	0.027	0.021	0.022	0.022	0.022	0.032	0.032
x25	0.565	0.042	0.026	0.026	0.027	0.029	0.028	0.029	0.022	0.022
x26	0.543	0.009	0.021	0.021	0.012	0.013	0.013	0.013	0.012	0.012
x27	0.535	-0.018	0.021	0.021	0.006	0.007	0.007	0.007	0.009	0.009
x28	0.644	0.081	0.034	0.035	0.045	0.047	0.047	0.047	0.040	0.041
x29	0.600	-0.032	0.022	0.023	0.007	0.008	0.008	0.008	0.014	0.014
x30	0.682	0.155	0.046	0.046	0.063	0.066	0.066	0.066	0.049	0.049
x31	0.744	0.164	0.046	0.046	0.058	0.062	0.061	0.062	0.049	0.049
x32	0.371	0.016	0.012	0.012	0.008	0.008	0.008	0.008	0.004	0.004
x33	0.366	0.030	0.014	0.014	0.018	0.019	0.019	0.019	0.014	0.014
x34	0.553	0.035	0.022	0.022	0.023	0.025	0.025	0.025	0.022	0.022
x35	0.628	-0.066	0.025	0.025	0.008	0.008	0.008	0.008	0.021	0.021
R^2	0.000	0.663	0.633	0.633	0.642	0.644	0.644	0.644	0.638	0.638
q				1.002		1.060		1.011		1.008

Table 2. Matrix of correlations between regression solutions

	r	**OLS**	**SVR**	**RR**	**ROLS**	**RNE**
r	1.000	0.396	0.909	0.718	0.718	0.890
OLS	0.396	1.000	0.676	0.884	0.884	0.622
SVR	0.909	0.676	1.000	0.911	0.911	0.955
RR	0.718	0.884	0.911	1.000	1.000	0.911
ROLS	0.718	0.884	0.911	1.000	1.000	0.911
RNE	0.890	0.622	0.955	0.911	0.911	1.000

Presented in Table 2 are the correlations between vectors of the main solutions from Table 1 (without the adjusted solutions proportional to the main solutions). All solutions, except OLS distorted by multicollinearity, are highly correlated with the vector **r**, with SVR beta-coefficients closest to the **r** structure and RNE next. OLS is close to ROLS and RR, while both are proportional so their correlation equals one. SVR has the highest relation with RNE, and both regularized solutions ROLS and RNE are highly connected too. The last two columns in Table 2 show that judging by the closeness to the vectors of **r** and SVR, the solution RNE outperforms the ROLS. It means that when the OLS is distorted by multicollineary impact, the regularization of normal equations RNE could be better than the regularization applied to the OLS criterion. However, the ROLS outperforms the RNE by the quality of fit R^2, as seen in Table 1.

Summary

Two new regression solutions non-prone to multicollinearity are considered by applying regularization of proportionality of the model coefficients to the pair correlations of the predictors with dependent variable. In one approach, this regularization is applied directly to the ordinary least squares objective (12), which leads to the system (14) with the solution (20). In another approach, such regularization is added to the relaxed system of normal equations (22), which yields the system (24) with solution (26). Comparison with ridge regression and Shapley value regressions, with an additional adjustment of the solutions to reach the best data fit, are considered as well. The developed techniques are presented in analytical form, and in contrast to ridge regressions they do not contain a free ridge parameter which in its turn could need an additional estimation. Both systems (14) and (24) correspond to robust non-prone to multicollinearity solutions for regression; they are simple and do not need any extensive iterative calculation. That is very important, especially for working with big data sets. Analytical and numerical results are very promising, and they show that the suggested methods can serve numerous practical needs of regression analysis.

References

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *The Annals of Statistics, 32*(2), 407-489. Retrieved from https://projecteuclid.org/euclid.aos/1083178935

Golub, G. H., Heath, M., & Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics, 21*(2), 215-223. doi: 10.1080/00401706.1979.10489751

Grapentine, A. (1997). Managing multicollinearity. *Marketing Research, 9*(3), 11-21.

Hansen, B. E. (2016). The risk of James-Stein and lasso shrinkage. *Econometric Reviews, 35*(8-10), 1456-1470. doi: 10.1080/07474938.2015.1092799

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer. doi: 10.1007/978-0-387-21606-5

Hawkins, D. M., & Yin, X. (2002). A faster algorithm for ridge regression of reduced rank data. *Computational Statistics & Data Analysis, 40*(2), 253-262. doi: 10.1016/s0167-9473(02)00034-8

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55-67. doi: 10.1080/00401706.1970.10488634

Hoerl, A. E., & Kennard R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 42*(1), 80-86. doi: 10.1080/00401706.2000.10485983

Lipovetsky, S. (2010). Enhanced ridge regressions. *Mathematical and Computer Modelling, 51*(5-6), 338–348. doi: 10.1016/j.mcm.2009.12.028

Lipovetsky, S. (2013). How good is best? Multivariate case of Ehrenberg-Weisberg analysis of residual errors in competing regressions. *Journal of Modern Applied Statistical Methods, 12*(2), 242-255. doi: 10.22237/jmasm/1383279180

Lipovetsky, S., & Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry, 17*(4), 319-330. doi: 10.1002/asmb.446

Lipovetsky, S., & Conklin, M. (2005). Ridge regression in two-parameter solution. *Applied Stochastic Models in Business and Industry 21*(6), 525-540. doi: 10.1002/asmb.603

Lipovetsky, S., & Conklin, M. (2015). Predictor relative importance and matching regression parameters. *Journal of Applied Statistics, 42*(5), 1017-1031. doi: 10.1080/02664763.2014.994480

Liu, X.-Q., & Gao, F. (2011). Linearized ridge regression estimator in linear regression. *Communications in Statistics – Theory and Methods, 40*(12), 2182–2192. doi: 10.1080/03610921003746693

Mason, C. H., & Perreault, W. D. (1991). Collinearity, power, and interpretation of multiple regression analysis. *Journal of Marketing Research, 28*(3), 268-280. doi: 10.2307/3172863

Roth, A. E. (Ed.). (1988). *The Shapley value: Essays in honor of Lloyd S. Shapley*. Cambridge, UK: Cambridge University Press. doi: 10.1017/cbo9780511528446

Shapley, L. S. (1953). A value for *n*-person games. In H. W. Kuhn & A. W. Tucker (Eds.), *Contributions to the theory of games* (Vol. II) (pp. 307-318). Princeton, NJ: Princeton University Press.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological), 58*(1), 267-288. Available from http://www.jstor.org/stable/2346178